

Is the “Law of the Jungle” Sustainable for the Internet?

Thomas Bonald
Orange Labs
Issy-les-Moulineaux, France
thomas.bonald@orange-ftgroup.com

Mathieu Feuillet*
Telecom ParisTech
Paris, France
mathieu.feUILlet@telecom-paristech.fr

Alexandre Proutiere
Microsoft Research
Cambridge, UK
alexandre.proutiere@microsoft.com

Abstract—In this paper we seek to characterize the behavior of the Internet in the absence of congestion control. More specifically, we assume all sources transmit at their maximum rate and recover from packet loss by the use of some ideal erasure coding scheme. We estimate the efficiency of resource utilization in terms of the maximum load the network can sustain, accounting for the random nature of traffic. Contrary to common belief, there is generally no congestion collapse. Efficiency remains higher than 90% for most network topologies as long as maximum source rates are less than link capacity by one or two orders of magnitude. Moreover, a simple fair drop policy enforcing fair sharing at flow level is sufficient to guarantee 100% efficiency in all cases.

I. INTRODUCTION

It is commonly accepted that the surprisingly high robustness of the Internet to the traffic growth of the last two decades is mainly due to the presence of congestion control. Since the introduction of congestion control algorithms in TCP in the late 1980s, we haven’t witnessed any congestion collapse similar to that of the young Internet. However, it is unclear whether the IETF will be able to enforce the use of TCP in the future. As a matter of fact, many applications already apply their own congestion control, if any, without complying with the “TCP-friendliness” principle [7], [9].

In this paper we seek to characterize the behavior of the Internet in the absence of any congestion control. As pointed out in [21], the induced packet losses may easily be recovered thanks to erasure coding. It would also be necessary to limit buffer sizes so as to maintain acceptable end-to-end delays. Under these assumptions, there is no clear reason why the rate of data transfers should be controlled. We thus go one step beyond the “decongestion control” scenario of Raghavan and Snoeren [21] and assume that all sources send data at their maximum rate without any control. Clearly, the network would then reach a totally different equilibrium point from today’s, with significant packet loss rates and the “Law of the Jungle” as a sharing principle. Most applications could adapt and survive. It remains to quantify the potential loss of efficiency in the utilization of network resources.

To this aim, we consider a traffic scenario with random arrivals of finite size flows and analyse the stability of the stochastic process formed by the number of flows in progress.

Flows are represented as fluid streams that thin on their path to the destination due to packet losses. Efficiency is assessed through the maximum traffic load the network can sustain under the stability condition. Surprisingly, there is generally no congestion collapse, even under usual FIFO scheduling and tail dropping. The worst scenario where all routers transmit packets that are eventually lost on their path to the destination occurs for cyclic networks only. In acyclic networks, our results show that efficiency remains generally higher than 90% as long as maximum source rates are less than link capacity by one or two orders of magnitude. Moreover, a simple fair drop policy enforcing fair sharing at flow level is sufficient to guarantee 100% efficiency in all cases.

While somewhat in contradiction with common belief that TCP plays a key role in the stability of the Internet, these results rely on the strong assumption that all applications use perfect erasure coding or selective retransmission schemes that guarantee the usefulness of any received packet. Another concern about the relevance of the considered scenario is the sensitivity of some key applications like voice and media streaming to high loss rates and varying available bandwidth. These issues are beyond the scope of the present paper, however, whose primary goal is to show that the absence of congestion control is not, *in principle*, an unviable option for the Internet.

The rest of the paper is structured as follows. Related work is presented in the next section. Section III describes the fluid model used to represent bandwidth sharing in the considered scenario without congestion control. The flow-level dynamics and the associated stability issues are introduced in Section IV. These results are used in Section V to estimate the loss of efficiency due to the absence of congestion control and validated through packet-level simulations in Section VI. Section VII concludes the paper.

II. RELATED WORK

Since the introduction of congestion control algorithms in TCP [10], a number of scheduling and dropping policies have been proposed to alleviate the problem of unresponsive flows, see [16], [18], [19], [22], [23], for instance. These policies have been designed with the aim of enforcing fair bandwidth sharing. In the presence of congestion control, they typically achieve max-min fairness [14]; we shall see that the absence

*Work done while in internship at Microsoft Research and Orange Labs.

of congestion control leads to a different allocation, that does not belong to the usual class of α -fairness [17].

The phenomenon of congestion collapse has rarely been considered under realistic traffic conditions with random flow arrivals. A notable exception is the paper by Massoulié and Roberts [15] that pointed out the potential instability of cycles. It was also shown in [2], [5], [24] that fairness stabilizes the network whenever possible, while the bias introduced by discriminatory scheduling policies may lead to loss of capacity. It turns out, however, that these critical flow-level dynamics are ignored in all papers that have previously considered performance in the absence of congestion control, see [1], [7], [12], [13], [21], [25] for instance.

III. BANDWIDTH SHARING

We first present a simple fluid model that captures the way flows may share bandwidth in the absence of congestion control. This model will be validated through packet-level simulations in Section VI.

A. Network model

Consider a network of L links. Denote by C_l the capacity of link l (in bit/s). A number of flows compete for access to these links. These flows are categorized into an arbitrary set of K classes. There are n_k class- k flows; each of these flows has an access rate to the network denoted by a_k and follows a route of length d_k in the network defined as an ordered set of distinct links $r_k = \{r_k(1), r_k(2), \dots, r_k(d_k)\}$.

Let φ_k be the throughput of each class- k flow. We refer to the vector $\varphi = (\varphi_1, \dots, \varphi_K)$ as the bandwidth allocation. It must satisfy the rate constraints:

$$\forall k, \varphi_k \leq a_k \quad \text{and} \quad \forall l, \sum_{k: l \in r_k} n_k \varphi_k \leq C_l. \quad (1)$$

A number of allocations have been considered in the literature to model the way flows actually share bandwidth in the Internet. These include max-min fairness, proportional fairness and α -fair policies [8], [11], [14], [17]. It turns out that, for all these allocations, the *order* in which class- k flows go through the set of links r_k does not matter. This is because flows are typically assumed to be rate-controlled, so that packet losses are neglected beyond the feedback they provide to the sources for rate adaptation. In particular, flows are generally considered as fluid streams and the rate of each class- k flow is assumed to be either equal to its access rate a_k or limited by the capacity of some link $l \in r_k$, independently of its position on route r_k ; there is no bandwidth waste, a property known as *Pareto efficiency*.

In the absence of congestion control, packet losses play a key role. Thus, while retaining the necessary abstraction of fluid streams, we need to describe the evolution of the rate of each flow going through the network. Specifically, we assume that each class- k flow initially transmits at full rate a_k and we denote by $\theta_k(i)$ its rate at the *output* of the i -th link on its route r_k , for all $i = 1, 2, \dots, d_k$. The actual throughput φ_k of

each class- k flow corresponds to its rate at the output of the last link on its route, $\theta_k(d_k)$.

Due to potential fluid loss at each link, we have:

$$\varphi_k = \theta_k(d_k) \leq \theta_k(d_k - 1) \leq \dots \leq \theta_k(1) \leq a_k. \quad (2)$$

Moreover, the total rate at the output of each link cannot exceed the capacity of this link, so that:

$$\forall l, \sum_{k: i: r_k(i)=l} n_k \theta_k(i) \leq C_l. \quad (3)$$

We say that a link is *saturated* if the corresponding inequality (3) is an equality. Note that the rate constraints (1) follow from (2) and (3). We shall see that, unlike usual allocations, the order in which class- k flows go through the set of links r_k does matter. Additionally, losses may result in bandwidth waste: the allocation is generally not Pareto efficient.

In order to characterize the allocation achieved in the absence of congestion control, it remains to determine how output rates depend on input rates at each link. This depends on the buffer management policy. In the considered fluid model, losses occur on saturated links only, when the total input rate exceeds capacity; the total output rate is then assumed to be equal to capacity. In the following, we consider two buffer management policies: tail dropping and fair dropping.

B. Tail dropping

The tail drop policy simply consists in dropping each incoming packet when the buffer is full. Assuming that all packets have the same probability of being dropped at the input of any given link, we deduce that, in the considered fluid model, the proportion of fluid lost is the same for all flows going through that link.

Specifically, let R_l be the total rate at the input of link l :

$$R_l = \sum_{k: r_k(1)=l} n_k a_k + \sum_{k: i: r_k(i+1)=l} n_k \theta_k(i).$$

The output rate of any class- k flow satisfies:

$$\theta_k(1) = a_k \min \left(\frac{C_l}{R_l}, 1 \right) \quad \text{if } l = r_k(1), \quad (4)$$

and for all $i = 2, 3, \dots, d_k$,

$$\theta_k(i) = \theta_k(i-1) \min \left(\frac{C_l}{R_l}, 1 \right) \quad \text{if } l = r_k(i). \quad (5)$$

The definition is non-ambiguous for *acyclic* networks, i.e., if links can be numbered in such a way that each route consists of an increasing sequence of link indices. The allocation then directly follows from applying (4) and (5) for all $l = 1, \dots, L$, successively. For general networks, the definition remains consistent in view of the following result, proved in the Appendix:

Theorem 1: The output rates obtained under tail dropping are uniquely defined by (4) and (5).

Consider the triangle network of Figure 1 for instance. This network has three links of unit capacity, and is shared by three

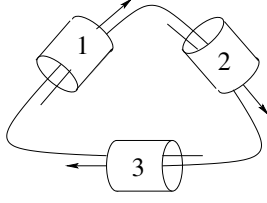


Fig. 1. A triangle network.

flows whose routes are $\{1, 2\}$, $\{2, 3\}$ and $\{3, 1\}$, respectively. All flows have the same access rate $a > 1/2$.

By symmetry, the common rate at the output of the first link of each route, denoted by $\theta(1)$, is the solution to the fixed point equation:

$$\theta(1) = \frac{a}{a + \theta(1)}.$$

We get $\theta(1) = \frac{1}{2}(\sqrt{a^2 + 4a} - a)$ and the rate at the output of the second link of each route $\theta(2) = \theta(1)/(a + \theta(1))$, that is:

$$\theta(2) = \frac{\sqrt{a^2 + 4a} - a}{\sqrt{a^2 + 4a} + a}. \quad (6)$$

This is the common throughput of all three flows. Note that the allocation is not Pareto-efficient since $\theta(2) < 1/2$. This bandwidth waste is due to the phenomenon of “dead” packets, a term coined by Kelly, Floyd and Shenker [12]: each flow uses some bandwidth of the first link of its route to transmit data that is lost at the second link.

C. Fair dropping

Under the fair drop policy, a packet belonging to that flow having the largest data volume in the buffer is dropped in case of buffer overflow. This is intended to guarantee to each incoming flow its fair share of bandwidth. In the considered fluid model, we shall assume that the output rates are max-min fair shares of the input rates at each link.

Specifically, let S_l be the fair share of any *saturated* link l . This is the unique rate such that the output rate of any class- k flow satisfies:

$$\theta_k(1) = \min(S_l, a_k) \quad \text{if } l = r_k(1), \quad (7)$$

and for all $i = 2, 3, \dots, d_k$,

$$\theta_k(i) = \min(S_l, \theta_k(i-1)) \quad \text{if } l = r_k(i), \quad (8)$$

and the total rate at the output of link l is equal to C_l . These rates follow from the following modified waterfilling procedure:

- (i) start from null rates;
- (ii) increase the rate of each flow at the same speed until some rate constraint (1) is reached;
- (iii) when some access rate constraint is reached, freeze the rate of the corresponding flows and keep increasing the rate of all other flows, if any; when the capacity of some link is reached, freeze the *downstream* rate of all flows going through this link and keep increasing the *upstream* rate of these flows and the rate of the other flows, if any;

(iv) stop when the rate of all flows is frozen.

This is similar to the standard waterfilling procedure leading to max-min fairness, except that one keeps increasing the upstream rate of those flows going through a saturated link. This defines a unique allocation, whose throughputs may be higher or lower than the max-min fair shares. Consider for example the network of Figure 2, consisting of three unit capacity links and four classes with a common unit access rate. For $n = (1, 1, 2, 3)$, we get under max-min fairness:

$$\varphi_1^{\text{mm}} = \frac{1}{2}, \quad \varphi_2^{\text{mm}} = \frac{1}{2}, \quad \varphi_3^{\text{mm}} = \frac{1}{5}, \quad \varphi_4^{\text{mm}} = \frac{1}{5},$$

while under fair dropping:

$$\varphi_1 = \frac{2}{3}, \quad \varphi_2 = \frac{1}{3}, \quad \varphi_3 = \frac{1}{5}, \quad \varphi_4 = \frac{1}{5}.$$

In the modified waterfilling procedure, keeping increasing the upstream rate of class 3 harms class 2 which, in turn, benefits to class 1.

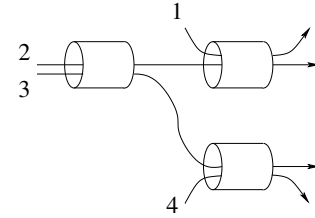


Fig. 2. A 3-link 4-class network.

The allocation φ obtained under fair dropping cannot be compared with max-min fairness but with the *min* allocation defined by:

$$\varphi_k^{\text{min}} = \min \left(a_k, \min_{l \in r_k} \frac{C_l}{\sum_{j: l \in r_j} n_j} \right). \quad (9)$$

For the example of Figure 2 with $n = (1, 1, 2, 3)$, we get:

$$\varphi_1^{\text{min}} = \frac{1}{2}, \quad \varphi_2^{\text{min}} = \frac{1}{3}, \quad \varphi_3^{\text{min}} = \frac{1}{5}, \quad \varphi_4^{\text{min}} = \frac{1}{5}.$$

The throughput of class- k flows under the min allocation follows from a waterfilling process after removing all links other than those of route r_k . It is easy to see that the resulting allocation is worse than that obtained with the modified waterfilling process. We deduce:

Lemma 1: Let φ be the allocation obtained under fair dropping. We have $\varphi_k \geq \varphi_k^{\text{min}}$ for all classes k .

This result will prove useful for the stability analysis of the allocation achieved under fair dropping.

IV. STABILITY ANALYSIS

In the rest of the paper, we consider a traffic scenario with random arrivals of finite size flows. We first analyse the stability of the stochastic process formed by the number of flows in progress. These results are used in the next section to estimate the ability of flows to effectively utilize network resources.

A. Flow-level dynamics

Let $n = (n_1, \dots, n_K)$ be the vector of the numbers of flows in progress, which we refer to as the network state. This is now a stochastic process, that evolves as new flows are generated by users and ongoing flows cease upon completion. We assume that class- k flows are generated according to a Poisson process of intensity λ_k and have independent, exponentially distributed sizes with mean $1/\mu_k$ (in bits). We define the traffic intensity of class- k flows as $\rho_k = \lambda_k/\mu_k$ (in bit/s). In any state n , we denote by $\varphi_k(n)$ the throughput of each class- k flow and by $\phi_k(n) = n_k \varphi_k(n)$ the total throughput of class- k flows.

Under the above assumptions, the network state defines a Markov process with transition rates λ_k from state n to state $n + e_k$ and $\phi_k(n)\mu_k$ from state n to state $n - e_k$, where e_k denotes the vector with 1 in the k -th component and 0 elsewhere. We say that the network is *stable* when this Markov process is ergodic, so that all flows have finite durations.

The usual stability condition is that traffic intensity is less than capacity for all links:

$$\forall l, \quad \sum_{k:l \in r_k} \rho_k < C_l. \quad (10)$$

This condition is known to be necessary for any allocation that satisfies the capacity constraints (1), see e.g. [3]. It is also sufficient for α -fair allocations, that typically represent the bandwidth sharing achieved by rate-controlled flows [2]. This result tends to show that congestion control stabilizes the network whenever possible.

In the following, we consider a network without any congestion control and derive the stability condition under tail dropping and fair dropping. The corresponding allocations are characterized by (4), (5) and (7), (8), respectively.

B. Tail dropping

It turns out that the usual stability condition (10) is generally not sufficient under tail dropping. We successively consider trees, cycles and lines to illustrate the sensitivity of the stability condition to the network topology.

1) *Downstream trees*: Downstream trees are defined as follows: (i) a common root link, say link 1, so that $r_k(1) = 1$ for all classes k ; (ii) any two classes j, k that have m links in common satisfy $r_j(i) = r_k(i)$ for all $i = 1, 2, \dots, m$. An example of downstream tree is presented in Figure 3.

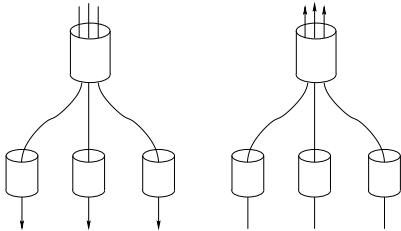


Fig. 3. A downstream tree (left), an upstream tree (right).

It may be easily verified that:

$$\forall k, \quad \phi_k(n) = \min \left(n_k a_k, \min_{l \in r_k} \left(C_l \frac{n_k a_k}{\sum_{j:l \in r_j} n_j a_j} \right) \right).$$

Thus the allocation coincides with weighted max-min fairness, with weights proportional to the access rates. This allocation is known to be stable under the usual stability condition (10) [5]. We deduce that the usual stability condition (10) is both necessary and sufficient for downstream trees.

2) *Upstream trees*: Upstream trees are defined as follows: (i) a common root link, say link 1, so that $r_k(d_k) = 1$ for all classes k ; (ii) any two classes j, k that have m links in common satisfy $r_j(d_j + 1 - i) = r_k(d_k + 1 - i)$ for all $i = 1, 2, \dots, m$. An example of upstream tree is presented in Figure 3.

We first consider a homogeneous upstream tree consisting of a unit capacity root link and K branches of capacity C . Each route goes through one branch and the root link. Traffic distribution is uniform; we denote by ρ the traffic intensity of each class.

Proposition 1: A homogeneous tree is stable under the usual stability condition $\rho < \min(C, 1/K)$.

Proof. The proof simply consists in observing that $\phi_k(n) \geq \min(C, 1/K)$ in all states n . Thus the network state is upper bounded by the state of K independent $M/M/1$ queues, each of load $\rho / \min(C, 1/K) < 1$. \square

Unlike homogeneous trees, the usual condition (10) may be not sufficient for the stability of heterogeneous trees. Consider a tree consisting of a unit capacity root link and $K - 1$ branches, each of capacity C . A route, corresponding to class 1, is directly connected to the root link, i.e., $d_1 = 1$; each of the $K - 1$ other routes goes through one branch and the root link. Classes $2, \dots, K$ have a common access rate a_2 and a common traffic intensity ρ_2 . The usual stability condition is:

$$\rho_1 + (K - 1)\rho_2 < 1, \quad \rho_2 < C. \quad (11)$$

Now let $\psi_1(n_1)$ be the minimum total rate of n_1 class-1 flows, obtained by letting the number of other flows grow to infinity:

$$\begin{aligned} \psi_1(n_1) &= \lim_{n_2, \dots, n_K \rightarrow \infty} \phi_1(n), \\ &= \min \left(n_1 a_1, \frac{n_1 a_1}{n_1 a_1 + (K - 1)C} \right). \end{aligned}$$

Define the corresponding total rate of any other class:

$$\begin{aligned} \psi_2(n_1) &= \lim_{n_2, \dots, n_K \rightarrow \infty} \phi_2(n), \\ &= \min \left(C, \frac{C}{n_1 a_1 + (K - 1)C} \right). \end{aligned}$$

Finally, consider the evolution of the number of class-1 flows, assuming these flows always get their minimum rate. Provided $\rho_1 < 1$, this is a stable birth-and-death process with stationary distribution:

$$\pi_1(n_1) = \pi_1(0) \frac{\rho_1^{n_1}}{\psi_1(1)\psi_1(2) \dots \psi_1(n_1)}.$$

Proposition 2: The considered heterogeneous upstream tree is stable if and only if $\rho_1 < 1$ and $\rho_2 < \kappa$, with:

$$\kappa = \sum_{n_1=0}^{\infty} \pi_1(n_1) \psi_2(n_1).$$

Proof. The result is a direct consequence of Theorem 2 and Theorem 3 of [4], since the allocation φ is partially decreasing in n and has an asymptotically uniform limit as n tends to infinity. \square

Note that, when $(K-1)C \geq 1$, the stability condition coincides with the usual one; we have in this case:

$$\begin{aligned} \kappa &= \sum_{n_1=0}^{\infty} \pi_1(n_1) \frac{C}{n_1 a_1 + (K-1)C}, \\ &= \sum_{n_1=0}^{\infty} \pi_1(n_1) \frac{1 - \psi_1(n_1)}{K-1}, \\ &= \frac{1 - \rho_1}{K-1}, \end{aligned} \quad (12)$$

from which (11) follows. When $(K-1)C < 1$, the stability condition is stricter than the usual one. In the limiting case $a_1 \rightarrow \infty$ for instance, class-1 flows have strict priority over other flows and the stability condition becomes $\rho_1 < 1$ and $\rho_2 < (1 - \rho_1)C$. The reduction of the stability region turns out to be slighter for low access rates, as illustrated by Figure 4 for $K = 2$ and $C = 0.5$.

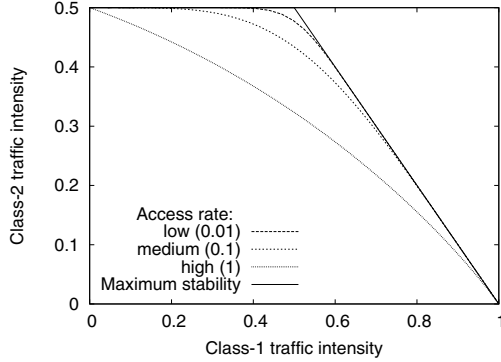


Fig. 4. Stability region of a heterogeneous upstream tree under tail dropping.

3) *Cycles:* We refer to a cycle as a network such that (i) for each class k and all $i < d_k$, $r_k(i+1) = r_k(i) + 1 \pmod{L}$; (ii) for all $l = 1, \dots, L$, there exists some k such that route r_k contains links l and $l+1 \pmod{L}$. The latter property ensures connectivity. The triangle of Figure 1 is an example of cycle.

Cycles are the worst network topologies for the tail drop policy, due to the phenomenon of “dead packets” described at the end of §III-B: whatever the initial condition, the network evolves towards a state with an infinite number of flows in progress, all links transmitting data that are eventually lost:

Proposition 3: A cycle is unstable whenever $\rho_k > 0$ for all classes k .

Proof. Let $C = \max_l C_l$, $a = \min_k a_k$ and $\rho = \min_k \rho_k$. Choose m such that:

$$\frac{C}{ma + C} C < \rho.$$

Let $\mathcal{K} \subset \{1, \dots, K\}$ be a minimal set such that the network restricted to classes \mathcal{K} is a cycle. Now assume that $n_k(t_0) \geq m$ for all $k \in \mathcal{K}$ at some time t_0 . Then there is some positive probability that $n_k(t_0) \geq m$ for all $k \in \mathcal{K}$ and all $t \geq t_0$. This is because the number of flows of each of these classes behaves as a transient single server queue with arrival rate λ_k and service rate:

$$\phi_k(t) \mu_k \leq \frac{C}{ma + C} C \mu_k < \lambda_k,$$

where the first inequality follows from the existence of some index $i \geq 2$ and some class $j \in \mathcal{K}$ such that $r_k(i) = r_j(1)$. The Markov process $n(t)$ is transient. \square

4) *Lines:* Lines are “broken” cycles. They are formally defined by the following two properties: (i) for each class k and for all $i = 1, 2, \dots, d_k - 1$, $r_k(i+1) = r_k(i) + 1$; (ii) for all $l = 1, \dots, L - 1$, there exists some k such that route r_k contains links l and $l+1$.

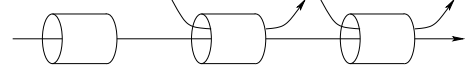


Fig. 5. A 3-link line.

An example of line is the following: there are $K = L$ flow classes with $r_k = \{k\}$ for all $k = 2, \dots, L$ and $r_1 = \{1, 2, \dots, L\}$. Such a line with $L = 3$ links is presented in Figure 5. In the limiting case where the access rates are infinite, the tail drop policy gives strict priority to short route flows, i.e., flows of classes $k = 2, \dots, L$. The network behaves as that considered in [2]. The stability condition is given by $\rho_k < C_k$ for all $k = 2, \dots, L$ and:

$$\rho_1 < \left(\min_l C_l \right) \times \prod_{k=2}^L \left(1 - \frac{\rho_k}{C_k} \right).$$

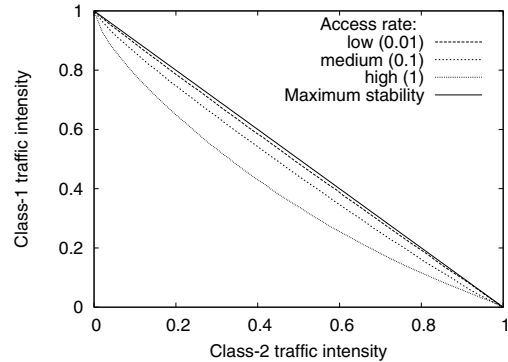


Fig. 6. Stability region of a homogeneous line under tail dropping.

Again, this condition is more constraining than the usual one, which is here $\rho_1 < C_1$ and $\rho_1 + \rho_k < C_k$ for all $k = 2, \dots, L$. As for upstream trees, the reduction of the stability region is actually marginal for low access rates. This is illustrated by Figure 6 for $L = 3$ links with $C_1 = \infty$, $C_2 = C_3 = 1$ and $\rho_2 = \rho_3$.

C. Fair dropping

We have seen that the absence of congestion control may reduce the stability region for some network topologies under tail dropping. A key result of the paper is that the fair drop policy stabilizes the network whenever possible:

Theorem 2: Under fair dropping, the network is stable under the usual stability condition (10).

Proof. Let $n^{\min}(t)$ be the Markov process associated with the min allocation described in §III-C. For any state n , denote by $\varphi(n)$ and $\varphi^{\min}(n)$ the vectors of per-flow rates under the fair dropping allocation and the min allocation, respectively. Denote by \leq the componentwise order on \mathbb{R}^K . We have for any two states n, n' such that $n \leq n'$:

$$\varphi(n) \geq \varphi^{\min}(n) \geq \varphi^{\min}(n'),$$

where the first and second inequalities follow from Lemma 1 and expression (9), respectively. By coupling, this implies that $n(t) \leq n^{\min}(t)$ a.s. for all time $t \geq 0$, provided $n(0) \leq n^{\min}(0)$. The proof then follows from the stability of the min allocation under the usual stability condition (10), cf. [6]. \square

V. RESOURCE UTILIZATION

We now apply the results of the previous section to quantify the loss of efficiency due to the absence of congestion control.

A. The Price of Anarchy

We refer to this loss of efficiency as the *price of anarchy*, a term that has been coined by Papadimitriou [20] to characterize the inefficiency of Nash equilibria in routing games. Here the game consists for each user in adapting their transmission rate; since each user has interest in increasing their transmission rate, the unique Nash equilibrium of our game is obtained when all users transmit at their respective maximum rates. The cost of this greedy strategy is a potential reduction of the stability region: flows are unable to effectively use the whole network capacity.

In the following, we define the price of anarchy P as the maximum reduction of traffic intensity imposed by the absence of congestion control. Specifically, we define:

$$P = \max_{u \in \mathbb{R}_+^K : \|u\|=1} \left(\frac{\alpha(u) - \beta(u)}{\alpha(u)} \right), \quad (13)$$

where $\|\cdot\|$ denotes the L^2 -norm on \mathbb{R}^K and $\alpha(u)$ and $\beta(u)$ denote the maximum values of α and β such that the vectors of traffic intensities $\alpha \times u$ and $\beta \times u$ satisfy the stability condition in the presence and in the absence of congestion control, respectively. Note that $0 \leq P \leq 1$. The price of anarchy depends on the network topology, the access rates and the considered buffer management policy.

B. Tail dropping

Under tail dropping, it follows from the previous results that the price of anarchy is equal to 0 for downstream trees and homogeneous upstream trees and equal to 1 for cycles. In general, P belongs to $(0, 1)$ and is very sensitive to the access rates. This is because the proportion of lost data strongly depends on the maximum source rates. Figures 4 and 6 suggest for instance that, for trees and lines, the price of anarchy vanishes when the access rates tend to 0. Proposition 4 below shows that this is indeed the case for the heterogeneous upstream trees considered in §IV-B. We believe that the result is in fact valid for any *acyclic* network, as defined in §III-B.

Proposition 4: Consider the heterogeneous upstream tree of §IV-B. The price of anarchy P tends to 0 as the class-1 access rate a_1 tends to 0.

Proof. In view of (12), it is sufficient to prove the result in the case $(K-1)C < 1$. Let m be such that:

$$\begin{aligned} \psi_1(n_1) &= n_1 a_1 \quad \text{if } n_1 \leq m, \\ \psi_1(n_1) &= \frac{n_1 a_1}{n_1 a_1 + (K-1)C} \quad \text{otherwise.} \end{aligned}$$

Define $p = \sum_{n_1=0}^m \pi_1(n_1)$. It may easily be verified that, as a_1 tends to 0, p tends to 1 if $\rho_1 < 1 - (K-1)C$ and to 0 otherwise. In the former case, we have:

$$\kappa \geq \sum_{n_1=0}^m \pi_1(n_1) \psi_2(n_1) = pC \rightarrow C,$$

while in the latter:

$$\begin{aligned} \kappa &\geq \sum_{n_1=m+1}^{\infty} \pi_1(n_1) \psi_2(n_1), \\ &\geq \sum_{n_1=0}^{\infty} \pi_1(n_1) \frac{1 - \psi_1(n_1)}{K-1} - p, \\ &= \frac{1 - \rho_1}{K-1} - p \rightarrow \frac{1 - \rho_1}{K-1}. \end{aligned}$$

In view of Proposition 2, the network becomes stable as a_1 tends to 0 provided the usual stability condition (11) is satisfied. \square

Figures 7 and 8 illustrate the decrease of the price of anarchy as a function of the assumed common access rate for the heterogeneous upstream tree of §IV-B2, with branch capacity $C = 1/K$, and the homogeneous line of §IV-B4, respectively. We observe that P is typically less than 0.1 for access rates less than 0.01, meaning that the efficiency in bandwidth utilization is higher than 90% in this practically interesting case.

Thus the loss of efficiency due to the absence of congestion control is significant in the presence of cycles only. It is worth observing, however, that the scenario described in §IV-B3 where all links transmit data that are eventually lost, leading to $P = 1$, is unlikely to occur in practice due to the presence of links at the input of the considered cycle. Consider for

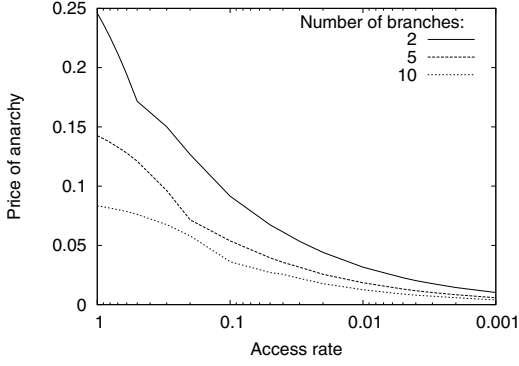


Fig. 7. Price of anarchy for a heterogeneous upstream tree under tail dropping.

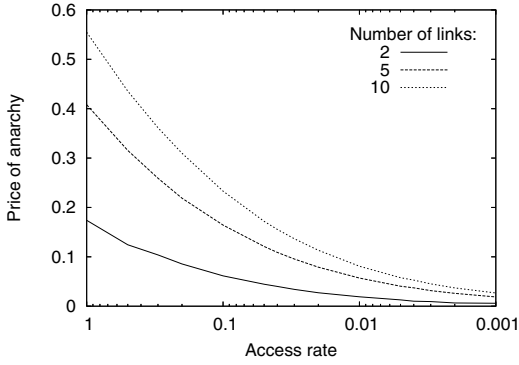


Fig. 8. Price of anarchy for a homogeneous line under tail dropping.

instance the triangle of Figure 9, with access links of capacity $A \geq 1/2$ (the three other links have unit capacity):

Proposition 5: The price of anarchy associated with the considered triangle is given by:

$$P = \frac{3 - \sqrt{1 + 4/A}}{1 + \sqrt{1 + 4/A}}.$$

Proof. It follows from (6) that, for a large number of flows of each class, each class is served at rate:

$$\varphi = \frac{\sqrt{1 + 4/A} - 1}{\sqrt{1 + 4/A} + 1}.$$

We deduce as in Propositions 1 and 3 that the stability condition is that the traffic intensity of each class is less than φ . The price of anarchy is then given by $P = 1 - 2\varphi$. \square

The impact of the access link capacity is illustrated by Figure 10: the price of anarchy decreases significantly with the access link capacity. This is also true for larger cycles, for which we use the following extension of Proposition 5. Consider a network consisting of $L = 2N$ links: N links forming a cycle, each of unit capacity, and N access links, each of capacity A . There $K = N(N - 1)$ classes, if we exclude the routes containing the N links of the cycle. We

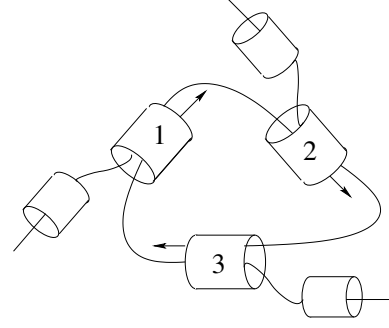


Fig. 9. A triangle with access links.

assume that $A \geq 1/d$, where $d = N - 1$ is the number of routes per link on the cycle. We have:

Proposition 6: The price of anarchy associated with the considered cycle with access links is given by $P = 1 - dA\delta^d$, where δ is the unique solution on $(0, A]$ to the equation:

$$\delta + \delta^2 + \dots + \delta^d = \frac{1}{A}. \quad (14)$$

Proof. First observe that the worst case is obtained when all routes have length N (the access link and d links on the cycle). As in the proof of Proposition 5, it remains to calculate the rate of each class, φ , in the presence of a large number of flows of each class. Denoting by $\theta(i)$ the rate of each class at the output of the i -th link on the cycle, we get:

$$\theta(i) = \theta(i - 1)\delta \text{ with } \delta = \frac{1}{A + \theta(1) + \dots + \theta(d - 1)},$$

using the convention $\theta(0) = A$. We deduce that $\theta(i) = A\delta^i$, from which (14) follows. The rate of each class is then given by $\varphi = \theta(d)$ and the price of anarchy by $P = 1 - d\varphi$. \square

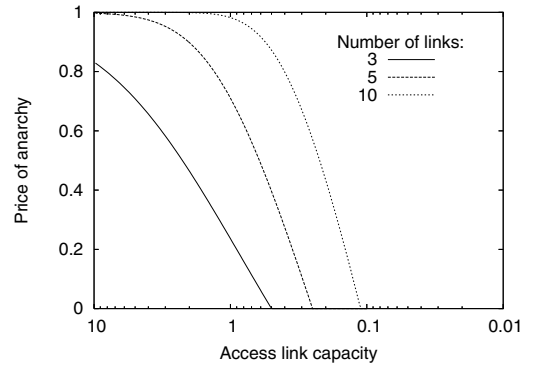


Fig. 10. Price of anarchy for a homogeneous cycle with access links under tail dropping.

C. Fair dropping

In view of Theorem 2, the price of anarchy associated with fair dropping is equal to 0. This simple buffer management policy is sufficient to guarantee the full utilization of network resources in all cases.

VI. PACKET-LEVEL SIMULATIONS

This section is devoted to the validation of the previous theoretical results through packet-level simulations.

A. Simulation setting

All packets have the same unit size. Flows are generated according to a Poisson process and have random sizes in number of packets, geometrically distributed with mean 100. The source associated with a class- k flow of size n sends packets at rate a_k along route r_k until the destination has received n packets; the source then immediately stops sending packets. There are 10^6 flow arrivals per simulation run; the network is declared stable if the average number of flows in each class is less than 500.

Each link is equipped with a buffer of 100 packets. To speed up the simulation, packets are served in random order instead of the FIFO discipline. Note that the fluid model described in Section III depends on the buffer management policy and *not* on the scheduling policy, as long as the latter is not biased against any class. Recall that the tail drop policy consists in dropping any incoming packet in case of buffer overflow, while the fair drop policy consists in dropping a packet from that flow having the largest number of packets in the buffer (in the presence of several such flows, one is chosen at random).

B. Results

First, we verified that under fair dropping, the network is stable under the usual stability condition for all topologies considered in Section IV: trees, cycles, lines. This is in line with Theorem 2. We also verified that, under tail dropping, downstream trees are stable under the usual stability condition whereas cycles are always unstable.

For those topologies where the price of anarchy is neither 0 nor 1, we obtain the results of Figures 11 and 12. The former shows the price of anarchy with respect to the number of classes for the heterogeneous upstream tree of §IV-B2 with branch capacity $C = 1/K$; the latter gives the price of anarchy with respect to the number of links for the homogeneous line of §IV-B4. In both cases, flows have a common access rate equal to 0.01, 0.1 or 1. Packet-level simulations are compared with flow-level simulations based on the allocation derived from the fluid model of Section III. We observe a good match of the results, showing that the complex packet-level dynamics are well captured by the simple fluid model we consider.

VII. CONCLUSION

Our results suggest that, somewhat surprisingly, the absence of congestion control does *not* lead to the well-known “tragedy of the commons”. The price of anarchy turns out to be moderate under tail dropping, at least for acyclic networks and low access rates, and null under fair dropping. Of course, these results are based on a number of simplifying assumptions whose impact should be quantified. In particular, we have assumed some perfect erasure coding scheme while all practical schemes involve some overhead.

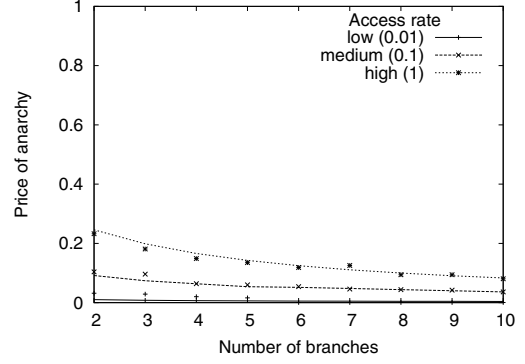


Fig. 11. Comparison of packet- and flow-level simulations for a tree.

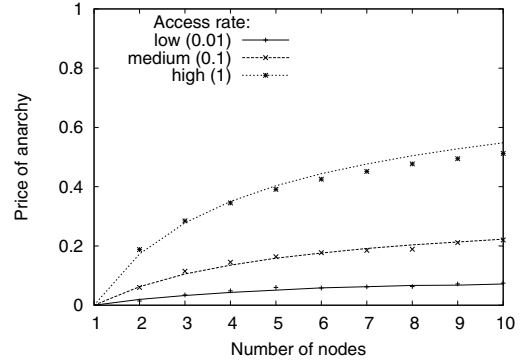


Fig. 12. Comparison of packet- and flow-level simulations for a line.

Other concerns about the practical relevance of the considered scenario include the phenomenon of local starvation induced by high-rate flows, that may increase the mean duration of low-rate flows, and the assumed ability of voice and media streaming applications to adapt to significant bandwidth variations. Again, these issues are alleviated by the fair drop policy, which tends to impose fair bandwidth sharing.

It is therefore not absurd to envisage the eventual abandon of congestion control, that may anyway be imposed by users and applications. This “ultimate” best-effort paradigm opens new perspectives regarding MAC layers, that may take advantage of the robustness of applications with respect to packet losses to further increase bandwidth utilization.

APPENDIX

PROOF OF THEOREM 1

Denote by $A_k = n_k a_k$ the total access rate of class- k flows. Without any loss of generality, we assume that $n_k > 0$ for all k , so that $A_k > 0$. Let $A = \max_k A_k$, $d = \max_k d_k$ and $N = \sum_{k=1}^K d_k$, where d_k is the route length of class k . Let $f : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ be the function defined by $y = f(x)$, with:

$$y_k(1) = A_k \min \left(\frac{C_l}{R_l}, 1 \right) \quad \text{if } l = r_k(1),$$

and for all $i = 2, 3, \dots, d_k$,

$$y_k(i) = x_k(i-1) \min\left(\frac{C_l}{R_l}, 1\right) \quad \text{if } l = r_k(i),$$

where

$$R_l = \sum_{k: r_k(1)=l} A_k + \sum_{k, i: r_k(i+1)=l} x_k(i).$$

We must prove that f has a unique fixed point, which then corresponds to the total output rate of each class.

Note that there exists some interval $[a, b] \subset (0, +\infty)$ such that $f([a, b]^N) \subset [a, b]^N$ and $f^d(\mathbb{R}_+^N) \subset [a, b]^N$. Since f is a continuous function on \mathbb{R}_+^N , it follows from Brouwer's theorem that f has at least one fixed point. In addition, any fixed point belongs to $[a, b]^N$.

To prove uniqueness, we need the following intermediate result. We denote by $\|\cdot\|$ the L^∞ -norm on \mathbb{R}^N .

Lemma 2: Let x^* be any fixed point of f . For any $r > 0$, denote by $B(r)$ the ball of center x^* and radius r . For any sufficiently small $r > 0$, there exists $\varepsilon > 0$ such that the function $f_\varepsilon = (1 - \varepsilon)f$ has a fixed point in $B(r)$.

Proof. Let R_l^* be the input rate of link l in state x^* :

$$R_l^* = \sum_{k: r_k(1)=l} A_k + \sum_{k, i: r_k(i+1)=l} x_k^*(i).$$

Without any loss of generality, we can assume that all links are saturated in state x^* in the sense that $R_l^* > C_l$ (otherwise, it is sufficient to restrict the analysis to non-saturated links). Using the fact that $x^* \in [a, b]^N$, we deduce that f is locally contracting in x^* : for all classes j, k and all indices $i = 1, 2, \dots, d_k - 1$ such that $l = r_j(1) = r_k(i+1)$, we have:

$$\left| \frac{\partial y_j(1)}{\partial x_k(i)} \right| = \frac{A_j C_l}{R_l^{*2}} \leq \frac{A_j}{R_l^*} < \frac{A_j}{A_j + a}.$$

Moreover, for all classes j, k and all indices $h = 2, 3, \dots, d_j$, $i = 1, 2, \dots, d_k - 1$ such that $l = r_j(h) = r_k(i+1)$,

$$\left| \frac{\partial y_j(h)}{\partial x_k(i)} \right| = C_l \frac{R_l^* - x_k^*(i)}{R_l^{*2}} \leq \frac{R_l^* - x_k^*(i)}{R_l^*} < \frac{N(A+b)}{a + N(A+b)}$$

if $j = k$ and $i = h - 1$, and

$$\left| \frac{\partial y_j(h)}{\partial x_k(i)} \right| = \frac{x_{h-1}^* C_l}{R_l^{*2}} \leq \frac{x_{h-1}^*}{R_l^*} < \frac{b}{a+b}$$

otherwise. Thus there exists some $\alpha < 1$ such that $f(B(r)) \subset B(\alpha r)$ for sufficiently small $r > 0$. For all $x \in B(r)$, we have:

$$\begin{aligned} \|f_\varepsilon(x) - x^*\| &= \|(1 - \varepsilon)f(x) - x^*\| \\ &\leq (1 - \varepsilon)\|f(x) - x^*\| + \varepsilon\|x^*\| \\ &\leq (1 - \varepsilon)\alpha r + \varepsilon\|x^*\|, \end{aligned}$$

which is less than r for sufficiently small $\varepsilon > 0$. We deduce that $f_\varepsilon(B(r)) \subset B(r)$, so that f_ε has some fixed point in $B(r)$. \square

To prove uniqueness, assume that f has two different fixed points, $x^{(1)}$ and $x^{(2)}$. In view of Lemma 2, there exist two non-overlapping balls $B^{(1)}$ and $B^{(2)}$ centered on $x^{(1)}$ and $x^{(2)}$, respectively, and some $\varepsilon > 0$ such that the function f_ε has at least one fixed point in each of these two balls. This is in contradiction with the fact that f_ε is a contracting function, which has a unique fixed point in \mathbb{R}_+^N . \square

REFERENCES

- [1] A. Akella, S. Seshan, R. Karp, S. Shenker, and C. Papadimitriou. Selfish behavior and stability of the internet: a game-theoretic analysis of tcp. In *SIGCOMM'02*, 2002.
- [2] T. Bonald and L. Massoulié. Impact of fairness on Internet performance. In *SIGMETRICS'01*, 2001.
- [3] T. Bonald, L. Massoulié, A. Proutière, and J. Virtamo. A queueing analysis of max-min fairness, proportional fairness, and balanced fairness. *Queueing Systems: Theory and Applications*, 53:65–84, 2006.
- [4] S. Borst, M. Jonckheere, and L. Leskela. Stability of parallel queueing systems with coupled service rates. *Discrete Event Dynamic Systems*, 2007.
- [5] G. de Veciana, T. Konstantopoulos, and T.-J. Lee. Stability and performance analysis of networks supporting elastic services. *IEEE/ACM Trans. Netw.*, 9(1):2–14, 2001.
- [6] G. Fayolle, A. de La Fortelle, J.-M. Lasgouttes, L. Massoulié, and J. Roberts. Best-effort networks: Modeling and performance analysis via large networks asymptotics. In *INFOCOM'01*, 2001.
- [7] S. Floyd and K. Fall. Promoting the use of end-to-end congestion control in the internet. *IEEE/ACM Transactions on Networking*, 7:458–472, 1999.
- [8] E. L. Hahne. Round-robin scheduling for max-min fairness in data networks. *IEEE J. Select. Areas Commun.*, 9:1024–1039, 1991.
- [9] M. Handley, S. Floyd, J. Padhye, and J. Widmer. TCP Friendly Rate Control (TFRC): Protocol Specification. RFC 3448, 2003.
- [10] V. Jacobson. Congestion avoidance and control. In *SIGCOMM'88*, 1988.
- [11] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.
- [12] T. Kelly, S. Floyd, and S. Shenker. Patterns of congestion collapse. *Unpublished*, 2003.
- [13] L. López, A. Fernández, and V. Cholvi. A game theoretic comparison of tcp and digital fountain based protocols. *Comput. Netw.*, 51(12):3413–3426, 2007.
- [14] L. Massoulié and J. Roberts. Bandwidth sharing: objectives and algorithms. *INFOCOM '99*, 3:1395–1403 vol.3, Mar 1999.
- [15] L. Massoulié and J. Roberts. Bandwidth sharing and admission control for elastic traffic. *Telecommunication Systems*, 15(1-2):185–201, 2000.
- [16] P. McKenney. Stochastic fairness queueing. In *INFOCOM'99*, 1999.
- [17] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.*, 8(5):556–567, 2000.
- [18] T. Ott, T. Lakshman, and L. Wong. SRED: stabilized RED. In *INFOCOM'99*, 1999.
- [19] R. Pan, B. Prabhakar, and K. Psounis. CHOCe - A stateless active queue management scheme for approximating fair bandwidth allocation. In *INFOCOM'00*, 2000.
- [20] C. Papadimitriou. Algorithms, games, and the Internet. In *STOC'01*, 2001.
- [21] B. Raghavan and A. Snoeren. Decongestion control. In *HOTNETS-V*, 2006.
- [22] M. Shreedhar and G. Varghese. Efficient fair queueing using deficit round robin. In *SIGCOMM'95*, 1995.
- [23] I. Stoica, S. Shenker, and H. Zhang. Core-stateless fair queueing: achieving approximately fair bandwidth allocations in high speed networks. In *SIGCOMM'98*, 1998.
- [24] H.-Q. Ye. Stability of data networks under an optimization-based bandwidth allocation. *IEEE Trans. Automatic Control*, 48(7):1238–1242, July 2003.
- [25] H. Zhang, D. Towsley, and W. Gong. TCP connection game: a study on the selfish behavior of TCP users. In *ICNP*, 2005.